EMbaRC

European Consortium
of Microbial Resource Centres

Grant agreement number: **228310**

Seventh Framework Programme

Capacities

Research Infrastructures

Combination of Collaborative Project and Coordination and Support Actions

Deliverable D.2.37
(formerly D.NA1.1.2)

Title: Guidelines for optimal formatting/annotation of data related to the biological materials

Due date of deliverable: M39

Actual date of submission: M42

Start date of the project: 1st February 2009 **Duration:** 44 months

Organisation name of the lead beneficiary: IP

Version of this document: V1.0

Dissemination level: PU

PU	**Public**	**X**
PP	Restricted to **other programme participants** (including the Commission)	
RE	Restricted to a **group defined by the Consortium** (including the Commission)	

EMbaRC is financially supported by the Seventh Framework Programme (2007-2013) of the European Communities, Research Infrastructures action

Document properties	
Project	EMbaRC
Workpackage	WP2 (NA1)
Deliverable	D.2.37 (formerly D.NA1.1.2)
Title	Guidelines for optimal formatting/annotation of data related to the biological materials
Version number	V1.0
Authors	David Smith and Chantal Bizet adapted from OECD and CABRI documents
Abstract	The GBRCN and EMbaRC consortia have updated this document to produce this standard which provides the basic requirements for data management and interoperability for BRCs. The Common Access to Biological Resources and Information project developed guidance for the operation of public service culture collections. This guidance was used as one of the building blocks for the OECD Best Practice Guidelines for Biological Resource Centres . The CABRI guidelines on catalogue production and date management have been adopted by many organisations and collections globally as a community standard. It also incorporates the key elements from the OECD working draft document, Best Practice Model for Minimum Requirements for Information Linkage and Exchange for all Biological Resource Centres and Global Biological Resource Centres Network formulated for the Debriefing Workshop on “Pilot Study on Operational Guidance for BRCs”, held on 14-15 November 2006 at the Belgian Science Policy, Wetenschapstraat 8 rue de la science, 1000 Brussels, Belgium. The various national, regional and global networks and project or initiative consortia of culture collections require information systems to provide access to their member collections resources. This document provides the guidance for culture collections and BRCs to manage and exchange their date to facilitate common practices and set a common standard for data management and interoperability. It is intended that one information system be created with portal sites for each community giving access to data on the biological resources maintained. This standard will provide best practice for the management of the information linkage and exchange for the GBRCN and EMbaRC (and other networks) as defined by the OECD in “Underpinning the future of life sciences and biotechnology”, OECD (2001). This document should be implemented in conjunction with “OECD General Guidelines for all BRCs”.
Validation process	Document prepared by IP in collaboration with CABI and submitted to the Executive Committee for agreement.

Revision table			
Date	***Version***	***Revised by***	***Main changes***
19 July 2012	1.0	Chantal Bizet	Included the mandatory audit of CABRI page 13 Included the WDCM list as a priority for collection acronyms Enter "B" for Bacteria or "A" for Archaea made clear in the MDS table page 35

Contents

Abbreviation key

BRC	Biological Resource Centre
BTWC	Biological and Toxin Weapons Convention
CABRI	Common Access to Biological Resources Information
CBD	Convention on Biological Diversity
EBRCN	European Biological Resource Centre Network
ECCO	European Culture Collection Organisation
EMbaRC	European Microbial Resources Consortium
EU	European Union
FDS	Full Data Set
GBIF	Global Biodiversity Information Facility
GBRCN	Global Biological Resource Centre Network
IDA	International Depositary Organisation
IP	Intellectual property
IPR	Intellectual Property Rights
ISO	International Standards Organisation
mBRC	microbial domain Biological Resource Centre
MDS	Minimum Data Set
MOSAICC	Microorganism Sustainable Use and Access Regulation International Code of Practice
OECD	Organisation for Economic Community Development
QMS	Quality Management System
RDS	Recommended Data Set
SOP	Standard Operating Procedure
UKNCC	UK National Culture Collection
WFCC	World Federation for Culture Collections
WHO	World Health Organisation

1 Background and Objectives

Work package NA1.1: *Harmonisation of methods and associated data* had the objectives:

- To establish common approaches and procedures for network activities so that all members of the consortium operate to the standards required to deliver high quality products and services to meet user expectations (Task NA1.1).
- Provide outputs delivering outreach to potential new network members and users to validate future customer base and business models (Task NA1.2)

The work package brought together three activities to better deliver high quality biological materials and information to the user. To provide a means of testing the best practice enabling an evidenced based selection of appropriate protocols for handling and storage of microorganisms and associated data. A second task aimed at delivering a harmonised mechanism for implementation and auditing the common quality system across the network i.e. this deliverable D.NA1.1.2 Guidelines for optimal formatting/annotation of data related to the biological materials. The third task specifically addressed a key aspect of the legal framework, biosecurity and specifically addresses risk assessment. Together they present a common harmonised and transparent operational framework, deliver the BRC standard, reduce user confusion and deliver them the high quality services they demand.

To open up the assets of biological resource centres to a larger user community a virtual catalogue that provides seamless access to all collection catalogues through a single portal interface is needed. This will be designed to be dynamic providing rapid synchronization of the information flow between the collection catalogues and the virtual catalogue. A standard information exchange format has been designed based on existing initiatives and proven effective in culture collections. Data types and existing formats were studied and the best way to go ahead was to adopt the CABRI catalogue production guidelines adapted to meet the OECD BRC Best Practice guidance[2] (OECD, 2007).

2 Guidelines for optimal formatting/annotation of data related to the biological materials

The Common Access to Biological Resources and Information project developed guidance for the operation of public service culture collections. This guidance was used as one of the building blocks for the OECD Best Practice Guidelines for Biological Resource Centres . The CABRI guidelines on catalogue production and date management have been adopted by many organisations and collections globally as a community standard. The GBRCN and EMbaRC consortia have updated this document to produce this standard which provides the basic requirements for data management and interoperability for BRCs. It also incorporates the key elements from the OECD working draft document, Best Practice Model for Minimum Requirements for Information Linkage and Exchange for all Biological Resource Centres and Global Biological Resource Centres Network formulated for the Debriefing Workshop on “Pilot Study on Operational Guidance for BRCs”, held on 14-15 November 2006 at the Belgian Science Policy, Wetenschapstraat 8 rue de la science, 1000 Brussels, Belgium.

The various national, regional and global networks and project or initiative consortia of culture collections require information systems to provide access to their member collections resources. This document provides the guidance for culture collections and BRCs to manage and exchange their date to facilitate common practices and set a common standard for data management and interoperability. It is intended that one information system be created with portal sites for each community giving access to data on the biological resources maintained. This standard will provide best practice for the management of the information linkage and exchange for the GBRCN and EMbaRC (and other networks) as defined by the OECD in “Underpinning the future of life sciences and biotechnology”, OECD (2001). This document should be implemented in conjunction with “OECD General Guidelines for all BRCs”.

This standard establishes a common basis for all culture collections and microbial domain Biological Resource Centres (m BRCs) globally. The CABRI quality standard is the result of reviewing database and catalogue information from all collection centres as well as their respective data authentication procedures.

Agreed data input rules and authentication procedures are presented in the "Data input & authentication" section.

The export format is the format required to enable the loading and indexing of the catalogues on CABRI host and it is described in details and by examples in the "Flat file preparation" section.

2.1.1 *Data input & authentication*

Data input and data authentication procedures have been designed and implemented by all collections in order to ensure the quality of information which is stored and used to describe each collection item. Different procedures have been designed, which are specific for the biological resource, these are: Fungi & Yeasts; Bacteria & Archaea; Animal cell lines; Hybridomas; Genomic libraries; Phages; Plasmids; Plant cell lines and Plant cell viruses. Although responsibility of the data is the depositor‘s responsibility, collection centres have to authenticate the data received from them. The authentication of data partly differs from centre to centre, but a CABRI label for quality assurance on data can be obtained when following these recommendations.

It is understood that each centre must be able to:

- provide traceability of data
- ensure back up and saves of data changes
- give signature for data entry
- give data entry typist level and profile.

To keep consistency for searching and retrieving from catalogues:

- each record must contain at least the minimum data set defined by CABRI for each type of collection
- spell checking for every field is a basic requirement
- vocabulary has to be checked against CABRI in-house reference lists
- arabic numbers preferred to Roman numbers except for the case in which the latter are included in a cell line name, virus name or any other proper noun (e.g. Amdur II and ASB XIV)
- language: English
- EMTREE thesaurus from EMBASE database produced by Elsevier is the CABRI reference thesaurus

Since mBRCs will use different platforms, it is necessary to adopt a standardised approach to certain scientific symbols. To avoid any errors due to incorrect reading of a character set, standard ASCII alternatives to symbols will be used. The alternatives include:

- Greek letters cannot be used, they must be fully spelled (write alpha, gamma, beta...)
- The ° symbol for temperature is to be omitted entirely (eg 37C replaces 37° C.
- No subscripts or superscripts are allowed (e.g. cm3 replaces cm3 and CO2 replaces CO2)

2.1.2 Networking of BRCs

(in the context of this document we refer to microbial and cell line culture collections or mBRCs and we use the acronym mBRC to identify them)

Any network of culture collections or BRCs should ensure standardisation of the information and data linkage requirements to enhance the efficiency of their activities and services. mBRCs have usually evolved independently to have different architectures in data systems, the mBRC system seeks to overarch information linkage and exchange of the information network using state-of-the-art techniques to manage heterogeneous collections of data.

Data linked and exchanged between BRCs and the clients may contain sensitive information both in terms of intellectual property and issues affecting data security regarding client information, human dignity and biosecurity. It is necessary to consider the means of authenticating a potential client who wishes to extract information and data from the system and to attach appropriate access facilities that match their status. In addition, the security of sensitive information must be ensured both during transit from one database to another and its security on arrival at the client database.

2.1.3 The Information Network[1]

The information network is a key function that supports information linkage and exchange among mBRCs and between mBRCs and clients, thereby assuring high quality and high efficiency, coupled with sufficient security, in accordance with the prescribed standards. The services of the mBRCs for information linkage and exchange are provided by its information network. The GBRCN and MIRRI are together developing a single information system based upon the implementation of the “CABRI standard”.

2.1.4 Roles and Services of the Information Network

The information network will underpin the information linkage between BRCs and their clients to:

Facilitate users’ access to biological resources which meet high quality standards by providing one-stop access to pertinent information and biomaterials for all countries

Provide linkages to associated databases in various disciplines (genetics, biochemistry, etc.)

Co-ordinate and combine catalogues and databases to meet the information requirements of biological science in the post-genomics era.

The information network aims to provide essential information and information tools to BRCs,

1. To provide essential information to mBRCs to avoid duplication of effort, broaden the coverage and share the load.
2. To help to assure information security for biological materials, data and information including those of relevance to biosecurity.
3. To provide networking of informatics tools for data analysis, comparison and visualisation
4. To promote the education of mBRC staff

---

[1] MIRRI – Microbial Resource Research Infrastructure – pan-European and GBRCN – Global Biological Resource Centre Network are working together toward such a network; www.gbrcn.org; www.mirri.eu

The information network aims to provide a platform for the biological resource communities.

1. Promote of international co-ordination, collaborative research and technology transfer
2. Provide example protocols and common approaches to enable mBRCs to comply with international rules, regulations and conventions where national guidance does not exist.
3. Raise awareness and the profile of BRC services through the portal site
4. To provide rapid access to expertise

3 Best practice for information linkage and exchange

This part of the guidance intends to describe the scope of best practice for information linkage and exchange for mBRCs and the mBRC information network. General guidelines for the informatics of mBRCs are prescribed in Guidance for the Operation of Biological Resource Centres (BRCs), General Guidelines for all BRCs[2].

3.1 Information Network[3]

The information network is the mechanism for the linkage and exchange of information and data. The information network is provided by the mBRC Network linking BRCs globally, which supports mBRCs and clients world-wide to communicate, link and exchange information and data.

3.2 The Portal Site of the mBRCs

A portal site should be implemented and managed by the Network which will provide a wide range of services for broader users and biological resource stakeholders.

mBRCs should be linked to the portal site. mBRCs may provide their data either direct through their own web links in addition to via the portal.

The information network provides a directory of members and draws strain data directly from the mBRC strain databases or via national or regional systems operating to the CABRI standard.

- mBRCs are registered in the directory and assigned a unique identifier. The institutional information can be either stored in the central directory or stored in mBRCs' server.
- The strain database is a virtual database that consists of databases distributed among mBRCs. This standard defines the standard data schema and protocol to make the distributed databases interoperable.

3.3 Execution of the Guidance for Information and Data Linkage Requirements

The Guidance should be reviewed periodically to revise any of its articles if any updating is required given the fact of the rapid progress of information and communication technologies (ICT)

3.4 Information Security

Whereas “Guidance for the Operation of Biological Resource Centres (BRCS), Biosecurity Principles” prescribes information security as an element of biosecurity, this guidance prescribes the requirements for information security at information linkage and exchange of general sensitive information to assure intellectual property rights, clients’ security, human dignity and biosecurity associated with the information and data linked and exchanged.

3.5 Future Development of the Guidance

The requirements for the information linkage and exchange for information network prescribed by this guidance will be further developed after the identification of the institutional architecture of the network secretariat is identified.

4 Scope

These guidelines provide the minimum requirements of information linkage and exchange for mBRCs. These guidelines could be applied to all domains of BRCs.

4.1 Definitions and Descriptions

4.1.1 Overall Structure

The information network should consist initially of a network of heterogeneous databases. Each of these databases should contain either information relating to strains held by a mBRC, other relevant data items or to composite data from several individual mBRCs.

It is important to bear in mind that a database relating to a particular mBRC is not necessarily located in the same place as the physical collection. For example, a database can contain information on UK micro-organisms but physically be located (hosted) in Paris. The important criterion is that the data is controlled by the responsible UK agency (the mBRC) using remote access.

The major feature of the information network should be to set up a Common Access System (CAS). The CAS should provide a transparent interface to the heterogeneous databases to provide seamless access to data.

It is expected that authentication of clients and their access rights will be dealt with at this level, thereby avoiding a multiplicity of requests for authorisation of users.

4.1.2 Unique Identifiers

Unique identifiers (UI) must be assigned to the member mBRCs and their holdings. There are existing systems such as Life Science Identifier (LSID), Digital Object Identifier (DOI), the accession number of sequence databases and WDCM number in the World Directory of Collections and Cultures of Microorganisms (the database of CCINFO).

The WDCM directory should be the reference point to the unique identifier of mBRCs. The unique identifier of the holdings should be secured by the Network directory and the management of holdings in each mBRC.

4.1.3 Information and Data (to Link and Exchange)

Information and data relate to the mBRCs concerned and their clients, to the biological materials they curate including their origin, and any other pertinent information accessible through the network information system.

4.1.4 Sensitive Information

Sensitive information is information that is for example of personal, private, confidential, business related that has potential for abuse and/or misuse.

NO SENSITIVE INFORMATION SHOULD BE TRANSFERRED TO THE NETWORK

4.1.5 Categories of Accession

Client access to information and data from mBRC's directly or through the information network should require the prior authentication of the client and should be limited in accordance with the status of clients. This can be implemented through a gate-guard function in the informatics system at the mBRC and the information network at the Network Secretariat.

4.1.6 Information network

The information network is the mechanism for the linkage and exchange of information and data. The information network should link mBRCs globally and should enable BRC and clients world-wide to communicate, link and exchange information and data.

4.1.7 Client

Clients are mBRCs which are linked to the Network and their users who have access to mBRCs either directly or through the information network services. Clients can deposit and use the biological material and/or the information and data relating to the material for research purposes or otherwise, provided they possess the necessary authorisation.

4.1.8 Client Authentication

If sensitive information is to be exchanged between a client and a mBRC, the individual clients should be authenticated and valid user-id and password should be provided by mBRCs.

5 Requirements

The data systems of mBRCs and Network should be robust, a proven and well-supported system

The informatics system employed by BRCs should provide appropriate facilities for information linkage and exchange of the mBRC. Corresponding parts of the OECD General Guidelines for All BRCs apply.

5.1 Requirements for the Information Network

Network services should provide an information environment that enables sharing of data elements, information and knowledge among the members of the Network, customers, supporters and the public.

A central core is required for the Network to activate, maintain and improve the information environment. The Secretariat curates a portal site that is an essential element of the information environment to be visible from mBRCs and beyond. A central core function will be further defined once the institutional architecture of the Network Secretariat is defined.

The portal site provides access to standards, helpdesk, announcement, news, communication tools, a virtual unified database of mBRCs, data analysis tools and links to other useful data sources. The tool for the portal site will be selected by the central core function and appropriate experts taking into account its cost and function.

5.2 Requirements for the Information Linkage and Exchange of mBRCs

The corresponding requirements for Informatics of mBRCs described in the OECD General Guidelines for All BRCs: apply with the following additions.

BRC should satisfy the CABRI standard for data processing to be become a Network member. The Network should add value to the resources provided by the member mBRCs by carrying out various service functions. To be a partner in CABRI requires a mandatory audit on all collection activities that are encompassed by the CABRI guidelines www.cabri.org.

5.2.1 Data Cleaning and Consistency

mBRCs and the Network

Data cleaning procedure should be used to detect errors in data and to then improve the quality and consistency. It is an essential part of information management and should be applied to the input of new data and applied to pre-existing information in current databases.

For the existing data, a series of checks should be carried out to ascertain their validity and completeness. As more BRCs become associated with the global network, more searches should be made for common classes of error to allow more efficient error correction.

For new data, wherever possible, inputting should be checked against authorised lists of not only scientific names but also thesaurus/ontology to prevent errors such as mistyping.

mBRCs
An mBRC should present the evidence of application of a recognised protocol appropriate for each data element (e. g. A comprehensive treatment of Data Cleaning can be found in Chapman, A.D. 2005. Principles and Methods of Data Cleaning – Primary Species and Species-Occurrence Data. Version 1.0 Report for GBIF, Copenhagen mainly in taxonomy The International Nucelotide Sequence Database (INSD) has prepared data submission tools that validate the data format and contents to some extent.)

Validation, Updating and availability of data as referred in the OECD Guidance for the Operation of Biological Resource Centres, General Guidelines for all BRCs.

5.2.2 Standardisation for the interoperability of information and data in mBRCs

mBRCs have maintained data management systems themselves. It is costly and not realistic to replace these legacy systems with a completely new system. It is feasible to make diverse data sources in mBRCs interoperable by ‘wrapping’ legacy systems. The CABRI standard defines the standards for this interoperability and it is the responsibility of the mBRC to implement such guidance.

The CABRI standard sets the standards for the data structure, the data representation and the data transportation taking into consideration existing standards for data processing, e.g. DarwinCore/DiGIR and ABCD schema/BioCASE for strain data, CCINFO for the organizational information of BRCs.

5.3 BRC services through Information Linkage and Exchange

5.3.1 Electronic Catalogue of Biological Resources

mBRCs should provide electronic catalogues of biological resources to clients through the facility of information linkage and exchange (e.g. web sites). mBRCs should restrict access to the electronic catalogues that include **4.4. Sensitive Information** to the authenticated clients in accordance with terms and conditions of deposit or any regulatory requirements (see 5.4.1 below)

5.3.2 Information Associated with Biological Materials; Data Sets

The following applies in conjunction with the corresponding Guidelines prescribed in the OECD General Guidelines for All BRCs and in domain-specific Guidelines.

mBRCs should provide Data Sets (Minimum Data Set, Recommended Data Set or Full Data Set – see below) to the users through information linkage and exchange when the biological resources are supplied to them and mBRCs must ensure the quality and consistency of Data Set during its transfer.

mBRCs should provide Data Sets to the users within the limits of ensuring the information security (biosecurity, protection of IPRs, clients' information and human dignity) of Data Set during its transfer and after its arrival at the user's site in accordance with prescribed authentication of users for the accession.

5.3.3 Data input & authentication for filamentous fungi and yeasts

Data Sets

- Minimum Data Set, General Definition
- Minimum data set for fungi and yeasts
- Recommended data set for fungi and yeasts
- Full data set

5.3.3.1 Mimimum Data Set: General definition

- The **Minimum Data Set** (MDS) is the set of data that is asked to describe **each product** from each collection. Each catalogue field has to be filled in (use a dash - if no information is given, when allowed).
- The MDS is required from new collection centres when they join CABRI.
- The **Recommended Data Set** (RDS): collection centres are also asked to supply data for recommended fields; however, if data are not available, the product is still acceptable
- Collection centres are free to provide data for other fields, which are part of the **Full Data Set.**
- A general MDS can be applied to each type of collections :

Identification	accession number, name
Origin	description, depositor, history..
Properties	
Culture conditions	medium, subculture routines...

Restrictions	
Hazard	
Bibliography	Reference paper, links to other projects

5.3.3.2 MDS for filamentous fungi and yeasts

Field	**Field label**	**Contents**
Strain number	Strain_number	Should contain: - collections’ own accession number of the strain
Other culture collection numbers	Other_collection_numbers	Should contain: - accession numbers of the strain in other collections. Cannot be given if the strain is unique in the collection.
Name	Name	Should contain the full scientific name of the strain: - Genus name and species epithet - Subspecies * - Variety * - Forma * - Forma specialis * - Authors of the name - *: only if applicable
Organism type	Organism_type	Should contain: - value "Fungi" or "Yeasts"
Restrictions	Restrictions	Should contain: - Release conditions/Restrictions - country-specific shipping restrictions - restrictions imposed by the depositor etc.
Status	Status	Should contain: - nomenclature status of the strain (like "type", "neotype", etc.) Does not apply to all strains.
History of deposit	History	Should contain the history of the element from isolator to depositor. - Name of isolator - Institute of isolator - Name of depositor - Institute of depositor - Names and institutes in between
Conditions for growth	Conditions_for_growth	Should contain: - Culture medium - Light conditions - Temperature conditions
Form of supply	Form_of_supply	Should contain: - form in which the strain will be sent to the customer.

5.3.3.2 RDS for filamentous fungi and yeasts

- MDS + the following fields:

Isolated from	Should contain name of substrate or host from which the strain was isolated.
Race	Should contain: - name of the race of the strain and authors of the race. Applies only to a small subset of the strains.
Geographic origin	Should contain name of country where strain originated from, followed by details on location.
Misapplied names	Should contain: - names incorrectly applied to the strain (misidentifications) that may still be in use. Does not apply to all strains.
Mutant	Should contain type of mutant if strain is a mutant strain. Applies only to a small number of strains.
Sexual state	Should contain sexual condition of the strain. Does not apply to all strains.
Literature	Should contain reference to publication in which the strain is described. Especially recommended for type strains.

5.3.3.3 Full Data Set for fungi and yeasts

- The **Minimum Data Set** (MDS) is the set of data that is asked to describe **each product** from each collection. Each catalogue field has to be filled in (use a dash - if no information is given). The MDS is required from new collection centres when they join CABRI. The **Recommended Data Set** : collection centres are also asked to supply data for recommended fields; however, if data are not available, the product is still acceptable
- Collection centres are free to provide data for other fields, which are part of the **Full Data Set.** Here are all the fields that are used by the collection centres:
- **Order and display fields**
- ***Set Field***
  MDS Strain Number
  MDS Other Collection Numbers
  MDS Name
  RDS Race
  MDS Organism Type
  MDS Restrictions
  MDS Status
  MDS History
  RDS Misapplied_names
  FDS Substrate
  RDS Geographic origin
  RDS Sexual state
  RDS Mutant
  FDS Genotype
  FDS Pathogenicity
  FDS Enzyme production
  FDS Metabolite production
  RDS Applications
  RDS Literature
  FDS Remarks
  MDS Conditions for growth

MDS Form of supply
FDS Price code

5.3.3.4 MDS for bacteria & archea

Field	**Field label**	**Contents**
Strain number	*Strain_number*	*Should contain:* *- collections' own accession number of the strain*
Other culture collection numbers	Other_collection_numbers	Should contain: - accession numbers of the strain in other collections. Can not be given if the strain is unique in the collection.
Restrictions	Restrictions	Should contain: - Release conditions/Restrictions - country-specific shipping restrictions - restrictions imposed by the depositor etc.
Organism type	Organism_type	Should contain: value "B" for bacteria or "A" for archaea.
Name	Name	Should contain the full scientific name of the strain: - Genus name and species epithet - Subspecies * - Pathovar * - Authors of the name *: only if applicable
Infrasubspecific names	Infrasubspecific_names	Should contain: - variety - designation - epithet - authors and reference Does not apply to all strains.
Status	Status	Should contain: - nomenclature status of the strain (like "type", "neotype", etc.) Does not apply to all strains.
History of deposit	History	Should contain the history of the element from depositor to isolator. - Name of depositor - Institute of depositor - Name of isolator - Institute of isolator - Names and institutes in between
Conditions for growth	Conditions_for_growth	Should contain: - Culture medium - Atmospheric and light conditions - Temperature conditions
Form of supply	Form_of_supply	Should contain: - the form in which the strain will be sent to the customer.

5.3.3.5 RDS for bacteria & archea

MDS + the following fields

Field	**Field label**	**Contents**
Serovar	Serovar	Should contain: the serovar name and author. Required for some medically important species. Does not apply to all strains.
Other names	Other_names	Should contain: names used for the strain in the past (not synonyms) that may still be in use. Does not apply to all strains.
Isolated from	Isolated_from	Should contain: name of substrate or host from which the strain was isolated.

Geographic origin	Geographic_origin	Should contain: name of country where strain originated from, followed by details on location.
Mutant	Mutant	Should contain: type of mutant if strain is a mutant strain.
Genotype	Genotype	Should contain: names of chromosomal markers of the strain. Especially recommended for strains of species for which many genetically modified strains exist. Does not apply to all strains.
Literature	Literature	Should contain: reference to publication in which the strain is described. Especially recommended for type strains.

5.3.3.6 FDS for bacteria & archaea

The **Minimum Data Set** (MDS) is the set of data that is asked to describe **each product** from each collection. Each catalogue field has to be filled in (use a dash - if no information is given). The MDS is required from new collection centres when they join CABRI.
The **Recommended Data Set**: collection centres are also asked to supply data for recommended fields; however, if data are not available, the product is still acceptable

Collection centres are free to provide data for other fields, which are part of the **Full Data Set**. Here are all the fields that are used by the collection centres:

Data Set	**Field Name**	**Description**
MDS	Strain number	Should contain collections' own accession number of the strain.
MDS	Other culture collection numbers	Should contain: - accession numbers of the strain in other collections. Can not be given if the strain is unique in the collection.
MDS	Restrictions	should contain: - Release conditions/Restrictions - country-specific shipping restrictions - restrictions imposed by the depositor etc.
MDS	Organism type	Should contain: bacteria or archaea.
MDS	Name	Should contain the full scientific name of the strain: - Genus name and species epithet - Subspecies * - Pathovar * - Authors of the name *: only if applicable
MDS	Infrasubspecific names	Should contain: - variety - designation - epithet - authors and reference Does not apply to all strains.
MDS	Status	Should contain: - nomenclature status of the strain (like "type", "neotype", etc.) Does not apply to all strains.
MDS	History of deposit	Should contain the history of the element from depositor to isolator. - Name of depositor - Institute of depositor - Name of isolator - Institute of isolator - Names and institutes in between

MDS	Conditions for growth	Should contain: - Culture medium - Atmospheric and light conditions - Temperature conditions
MDS	Form of supply	Should contain: - the form in which the strain will be sent to the customer.
RDS	Serovar	Should contain: - the serovar name and author. Required for some medically important species. Does not apply to all strains.
RDS	Other names	Should contain: - names used for the strain in the past (not synonyms) that may still be in use. Does not apply to all strains.
RDS	Isolated from	Should contain name of substrate or host from which the strain was isolated.
RDS	Geographic origin	Should contain name of country where strain originated from, followed by details on location.
FDS	Sexual state	
RDS	Mutant	Should contain type of mutant if strain is a mutant strain.
RDS	Genotype	Should contain names of chromosomal markers of the strain. Especially recommended for strains of species for which many genetically modified strains exist. Does not apply to all strains.
RDS	Literature	Should contain reference to publication in which the strain is described. Especially recommended for type strains.
FDS	Pathogenicity	
FDS	Enzyme production	
FDS	Production of metabolites	
FDS	Applications	
FDS	Catalogue entry	
FDS	Remarks	
FDS	Price code	
FDS	Plasmids	

5.3.4 Data sets for animal and human cell lines

5.3.4.1 MDS for animal & human cell lines

Field	**Field label**	**Contents**
Cell line name	Cell_line_name	should contain: - name of the cell line as given by the depositor
Accession number	Accession_number	should contain the collections' own accession No or cell code
Brief description	Brief_description	should contain ["brief" listing]: - species

		- strain - tissue - tumour - pathology - transformed/ transfected
Description and other remarks on the cell line	Description	should contain ["verbose" listing]: - species - strain - tissue - tumour - pathology - transformed/ transfected - age - sex - derived from
Depositor name and address	Depositor	should contain: - depositor name - address of the Institution
Reference paper	Bibliographic_references	should contain: - the original paper [if available]
Morphology and growth characteristics	Morphology	should contain: - morphology - growth characteristics (monolayer, suspension, etc.)
Culture conditions	Culture_conditions	should contain: - culture medium - subculture routine - incubation - harvest
Viruses	Viruses	should contain: - absence/presence - type of virus
Properties	Properties	should contain: - immunology - receptors - susceptibility to viruses
Release conditions/ Restrictions	Release_conditions	should contain: - release conditions/restrictions - patents
Hazard	Hazard	should contain: - hazard classification of the cell line [e.g. class II containment] - EU related hazard information rules - country-specific related hazard information rules

5.3.4.2 RDS for animal cell lines: MDS + passage N° + species validation tests

Field	**Field label**	**Contents**
Passage	Passage_number	should contain: passage number
Species validation tests	Species_validation	should contain: species validation tests carried out to confirm species of cell line

5.3.5 Data sets for plant cell lines

5.3.5.1 MDS for plant cell lines

Field name	**Description**
Scientific name	should contain: - botanical name of initial plant
Family	should contain: - family name of the species
Accession number	should contain: - collections' own accession Number
Strain designation	should contain: - designations characterising the cell line, usually given by depositor or initiator
Depositor name and address	should contain: - depositor name - address of the Institution
Culture conditions	should contain: - name of main cultivation medium (abbreviated) - recommended growth temperature (°C) - Humidity of growth chamber (%) - Light conditions (lux) - Light regime (hours)
Release conditions/ Restrictions	should contain: - Release conditions/Restrictions - country-specific shipping restrictions - restrictions imposed by the depositor etc.
Hazard/BioHazard	should contain: - hazard classification of the plant cell - EU related hazard information rules - country-specific related hazard information rules
Additional information	should contain: - any additional information about the cell line
History	Should contain: - history of the cell line before arrival at the resource centre
Date of initiation	date of culture initiation

5.3.5.2 *Full Data Set for plant cell lines*

- The **Minimum Data Set** (MDS) is the set of data that is asked to describe **each product** from each collection. Each catalogue field has to be filled in (use a dash - when no information is given). The MDS is required from new collection centres when they join CABRI.
- The **Recommended Data Set** : collection centres are also asked to supply data for recommended fields; however, if data are not available, the product is still acceptable
- Resources centres are free to provide data for other fields, which are part of the **Full Data Set** (FDS)
- If and only if resource centres plan to put in their own FDS an information that is already present in other FDSs THEN they can try to adhere to existing fields:

Data Set	**Field name**	**Description**
MDS	**Accession number**	should contain: - collections' own accession Number
MDS	**Scientific name**	should contain: - botanical name of initial plant
MDS	**Family**	should contain: - family name of the species
MDS	**Strain designation**	should contain: - designations characterising the cell line, usually given by depositor or initiator
MDS	**Depositor name and address**	should contain: - depositor name - address of the Institution
MDS	**Culture conditions**	should contain: - name of main cultivation medium (abbreviated) - recommended growth temperature (°C) - Humidity of growth chamber (%) - Light conditions (lux) - Light regime (hours)
MDS	**Release conditions/ Restrictions**	should contain: - Release conditions/Restrictions - country-specific shipping restrictions - restrictions imposed by the depositor etc.
MDS	**Hazard/BioHazard**	should contain: - hazard classification of the plant cell - EU related hazard information rules - country-specific related hazard information rules
MDS	**Additional information**	should contain: - any additional information about the cell line
MDS	**History**	Should contain: - history of the cell line before arrival at the resource centre
MDS	**Date of initiation**	date of culture initiation

FDS	**Initiation material**	
FDS	**Date of accession**	

5.3.5.3 Data sets for phages

5.3.5.3.1 MDS for phages

Field	**Field label**	**Contents**
Collection number	Collection_number	Contains the number of the element in the collection under which the element is distributed.
Element name	Name	Contains the name of the element.
Element type	Type	Contains the type of the element
Other culture collection numbers	Other_collection_numbers	Number given by other collections to the phage
Restricted distribution	Restricted_distribution	Should contain: • release conditions/restrictions • country-specific shipping restrictions • restrictions imposed by the depositor etc.
Literature	Literature	Original publication in which the element is described.
History of deposit	History_of_deposit	Contains the history of the element from author to depositor.
Host for propagation	Host_for_propagation	Description of suitable host strains (on family, genus, species level), where possible completed with requirements for infection on phenotypic level (e.g. gram-negative bacteria, or *E. coli* LamB+)
Host used for propagation	Host_used_for_propagation	Name of the host strain (species and element name/collection number) that was used by the collection for propagation of the phage
Lysogenicity	Lysogenicity	Lysogenicity of the phage. Logical field: "yes" or "no".
Virus used for	Virus_used_for	Possible applications of the phage

5.3.5.4 RDS for phages

MDS + cell surface receptor

Cell receptor surface	Name of the cell surface receptor

5.3.5.4.1 Full data set

FDS for phages

Data Sets

The **Minimum Data Set** (MDS) is the set of data that is asked to describe **each product** from each collection. Each catalogue field has to be filled in (use a dash - if no information is given).The MDS is required from new collection centres when they join CABRI. The **Recommended Data Set**: collection centres are also asked to supply data for recommended fields; however, if data are not available, the product is still acceptable

Collection centres are free to provide data for other fields, which are part of the **Full Data Set.** Here are all the fields that NCCB & DSMZ are using:

Data Set	**Field**	**Description**
MDS	Collection number	Contains the number of the element in the collection under which the element is distributed.
MDS	Element name	Contains the name of the element.
MDS	Element type	Contains the type of the element
MDS	Other culture collection numbers	Number given by other collections to the phage
MDS	Restricted distribution	should contain: • Release conditions/Restrictions • country-specific shipping restrictions • restrictions imposed by the depositor etc.
MDS	Literature	Original publication in which the element is described.
MDS	History of deposit	Contains the history of the element from author to depositor.
MDS	Host for propagation	Description of suitable host strains (on family, genus, species level), where possible completed with requirements for infection on phenotypic level (e.g. gram-negative bacteria, or *E. coli* LamB+)
MDS	Host used for propagation	Name of the host strain (species and element name/collection number) that was used by the collection for propagation of the phage
MDS	Lysogenicity	Lysogenicity o the phage, Logical field: "yes" or "no".
MDS	Virus used for	Possible applications of the phage
RDS	Cell surface receptor	Name of the cell surface receptor
FDS	Medium	Medium

5.3.5.5 Data sets for plasmids

MDS for plasmids

Field	**Field label**	**Contents**
Collection number	Collection_number	Number of the element in the collection under which it is distributed

Name	Name	Name of the element
Other culture collection numbers(*)(**)	Other_culture_collection_numbers	Numbers of the element under which it is distributed by other culture collections
Type	Type	Element type
Class	Class	Origin of the element: non-recombinant or recombinant
Literature(**)	Literature	Original publication in which the element is described.
History of deposit(**)	History_of_deposit	History of the element from author to depositor, in reverse chronology (depositor < ... < author)
Restricted distribution	Restricted_distribution	Restrictions for distribution of the element (hazard group, “for research purposes only”, patent info, import/export regulations, etc...).
Host for distribution	Host_for_distribution	Taxonomic designation and strain designation of the host organism in which the element is distributed.
Medium	Medium	Name or collection number of the medium on which the host organism is preferably grown.
Selectable phenotype	Selectable_phenotype	Feature encoded by the element enforcing its host strain to maintain the element in order to survive under appropriate cultivation conditions.
Replicon(**)	Replicon	Names of replicon elements.
Host range(**)	Host_range	Hosts in which the element can be propagated and/or used and, when applicable, the peculiar features that are necessary for making use of the special characteristics of the element.
Properties and applications	Properties_and_applications	Description of the properties of the element, as well as its application possibilities.

(*) If the content of the field 'Other culture collection numbers' cannot automatically be generated at the main site it will be removed from the MDS for plasmids.

(**)This field does not apply to all strains, you may use the dash to express it.

Full data set for plasmids

5.3.5.6 Data sets for viruses

5.3.5.6.1 MDS for viruses

Field Name	**Description**
Accession Number	contains: - Unique identifier
Virus Name	contains: [verbose listing]: - Species Name
Virus Name Abbreviation	contains: - Species name abbreviation

Former Name	may contain: - Former name of the isolate, e.g., in case a patho- or serotype has been assigned to another species.
Genus	contains: - Genus to which the virus species has been assigned
Pathotype, Serotype, Strain	should contain: - "Pathotype: ", "Serotype: " or "Strain: " followed by pathotype, serotype or strain designation.
Original Host	should contain: - Original host plant: species (latin binominal), - subspecies (latin), - variety (latin), cultivar or - breeding line (original designation), - vernacular name
Geographic Origin	should contain: - Country, and, if available, more detailed information on the geographic position where the original host plant was found.
Isolate History	should contain: - Name of the scientist, who isolated the specimen and of persons or organisations that kept it before accessioning.
Reference Isolate	should contain: - Literature on the virus isolate
Quarantine Regulations	should contain: - Known EU quarantine regulations for the virus species
Remarks	Special features of the isolate, e.g., virulence

5.3.5.6.2 Full Data Set for viruses

For viruses the FDS is equal to the MDS.

5.3.5.7 *Data sets for genomic libraries*

5.3.5.7.1 *Minimum data set for genomic libraries*

MDS for genomic libraries

Field	**Field label**	**Contents**
Library Name	Library_name	Should contain Library Identifier
Organism	Organism	Should contain the organism represented in the Library
Type	Type	Should contain the type of Library
Vector	Vector	Should contain the vector name, more specific than type as will be specific about vector
Insert Size	Insert_size	Should contain average size of inserts in vector in Kb
Library coverage	Coverage	Should contain the degree of coverage
MTA required	MTA	Should contain a Material Transfer Agreement, if required

5.3.5.8 *Data sets for hybridomas*

5.3.5.8.1 *MDS for hybridomas*

Field name	**Description**
Cell line [Hybridoma] name	should contain: - Name of the hybridoma as given by the depositor
Accession number	should contain: - the collections’ own accession number
Brief description	should contain ["brief" listing]: - species - strain - antibody specificity
Description and other remarks on the hybridoma	should contain ["verbose" listing]: - species - strain - antibody specificity - donor [immunised animal] - antibody produced - cross-reaction - fusion partner
Depositor name and address	should contain: - depositor name - address of the Institution
Reference paper	should contain: - the original paper [if available]

Morphology and growth characteristics	should contain: - morphology - growth characteristics (monolayer, suspension, etc.)
Culture conditions	should contain: - culture medium - subculture routine - incubation - harvest
Release conditions/ Restrictions	should contain: - release conditions/restrictions - patents
Hazard	should contain: - hazard classification of the hybridoma - EU related hazard information rules - country-specific related hazard information rules

Full data set

5.4 Information Security

Information security of information linkage and exchange is the security of ***Sensitive Information*** defined at **4.4.** during information linkage and exchange directly or through the information network including those after arrival of information at different databases (clients')

The data are owned by the mBRCs. However, it is to be understood that the data in the Network standard data set are in the public domain.

AN ALTERNATIVE IS THAT NO SENSITIVE INFORMATION SHOULD BE TRANSFERRED TO THE NETWORK

5.4.1 Information Security Measures

The Network and mBRCs should implement appropriate measures (protocols, tools and standards) at their own information networks and informatics system respectively to assure reasonable information security of client information, data exchanged and data associated with biological materials supplied or deposited for protection against unauthorised accession. There are existing systems, *e.g.* authentication by user-id and password, encryption, encryption of messages and restriction of IP addresses.

5.4.2 Prescription for Sensitive Information

The mBRC Manager or an assigned responsible person should determine ***Sensitive Information*** which includes:

1. Client information
2. Information derived from donors of human-derived biological resources (materials and information)
3. Information and data associated with intellectual property rights
4. Information which involves biosecurity risks
5. Confidential information regarding the mBRC and its staff and
6. Any other information which should be secured from unauthorised access.

5.4.3 Information Security for Biosecurity

Regarding the risk assessment for, and categorisation of, accession to the information, which involves biosecurity risks Guidance for the Operation of Biological Resource Centres (BRCs), Biosecurity Principles should apply (or that subsequently provided by the OECD Biosecurity Task Force 2010-2011).

5.4.4 Management of Information Security

The mBRC Manager or an assigned responsible person should determine appropriate ***Categories of Accession*** defined at **4.5.** to ***Sensitive Information*** in accordance with the status of clients and the nature of the ***Sensitive Information***.

Information security measures of mBRCs should provide gate-guard function for information linkage and exchange. Information security measures of mBRCs should execute user-access control in accordance with prescribed ***Categories of Accession*** and ***Client Authentication*** defined at **4.8.** (using valid user-id and password) to assure appropriate accession to and protection for ***Sensitive Information*** linked and exchanged.

Information security measures of the Network should provide gate-guard function for information network. Information security measures of the Network should execute user-access control in accordance with ***Categories of Accession*** determined by the mBRC and ***Client Authentication*** to assure appropriate accession to and protection of the ***Sensitive Information*** that is linked and exchanged through the information network.

ALTERNATIVELY NO SENSITIVE INFORMATION SHOULD BE TRANSFERRED TO THE NETWORK

6 Management of Information Linkage and Exchange

6.1 Tools, Protocols and Schemes

An mBRC should update its facilities and tools for information linkage and exchange, where appropriate and available, in timely fashion to meet new demands for mBRC services and functions respecting of the rapid progress of information communication technologies.

6.2 The Staff assigned for the Managements and Services

mBRCs should assign appropriate staff for the following managements, maintenance and services:

- The maintenance of the facilities and tools for information linkage and exchange
- The management of and services through the portal site of the mBRC
- The services for Clients accession through information linkage and exchange
- The management of data security at information linkage and exchange and the maintenance of the measures and tools

Training courses should be organized to make mBRC staff more informatics literate.

6.3 Data entry and validation

6.3.1 Data entry and validation for bacteria and fungi

Field	Strain number
Description	collections' own accession no of the strain, consists of collection acronym followed by a number separated with a blank.
Input process	Make sure unique number is given to each strain
Required for	MDS

Other culture collection numbers

Field	Other culture collection numbers
Description	accession numbers of the strain in other collections. Numbers are separated by semicolon. Can not be given if the strain is unique in the collection.

Input process	**New strain**: as given by depositor **Existing strain:** if strain is sent to another collection, get and add accession number used in that collection Collection acronyms: CABRI agreed list Numbers are separated by semicolon
Required for	MDS

Restrictions

Field	Restrictions
Description	value "no", or a text field describing the appropriate restrictions: hazard group, import/export regulations, .. The text may contain codes which refer to text files explaining the restrictions in more detail.
Input process	Enter ‘No’, or actual restrictions, which may be linked to a file for conditions of delivery
Required for	MDS

Organism type

Field	Organism type
Description	"Fungi" for filamentous fungi or "Yeasts" for yeasts.
Input process	Enter ‘Fungi’ or ‘Yeasts’
Required for	MDS

Name

Field	Name
Description	full scientific name of the strain: Genus name and species epithet Subspecies * Variety * Forma * Forma specialis * Authors of the name *: only if applicable
Input process	Enter full scientific name as given by depositor and confirmed (or changed) by collection. Names of authors of the name are included. CABRI list of names is available, but not complete
Required for	MDS

Race

Field	Race
Description	name of the race of the strain and authors of the race. Applies only to a small subset of the strains
Input process	Enter name or number of the race as given by depositor and confirmed (or changed) by collection. Names of authors of the race are included after a comma.
Required for	RDS

Status

Field	Status
Description	nomenclatural status of the strain (like "type", "neotype", etc.) Does not apply to all strains
Input process	Enter information on type ("type", "neotype", etc.). In case of synonyms (status refers to another taxon than the one mentioned in the NAME field), add 'of' and the scientific name of the synonym Format and reference list: see MINE format (Gams et al., J Gen Microbiol 1988;134:1667-1689).

Required for	MDS

Misapplied names

Field	Misapplied names
Description	names incorrectly applied to the strain (misidentifications) that may still be in use. Does not apply to all strains
Input process	Enter previous name of strain when the strain is reidentified, and the old name is not a synonym of the new name. Same format and reference list as Name field.
Required for	RDS

History of deposit

Field	History
Description	history of the element from isolator to depositor. Name of isolator Institute of isolator Name of depositor Institute of depositor Names and institutes in between
Input process	Enter information as given by depositor Format: see MINE format (Gams et al., J Gen Microbiol 1988;134:1667-1689).
Required for	MDS

Conditions for growth

Field	Conditions for growth
Description	Culture medium Light conditions Temperature conditions
Input process	Each collection has to provide its own list of culture media and recipes, medium provided for strain is linked to medium in this file Atmospheric conditions should only be given if they are special. Additional remarks on the cultivation like 'freshly prepared medium is necessary' or 'extended incubation time…' if necessary'
Required for	MDS

Form of supply

Field	Form of supply
Description	form in which the strain will be sent to the customer
Input process	Values: ‘Active', ‘Dried', 'Dried ice'
Required for	MDS

Isolated from

Field	Isolated from
Description	name of substrate or host from which the strain was isolated.
Input process	Enter information as given by depositor Format: see MINE format (Gams et al., J Gen Microbiol 1988;134:1667-1689). Reference list: not available, could be compiled from data.
Required for	RDS

Geographic origin

Field	Geographic origin
Description	name of country where strain originated from, followed by details on location
Input process	Enter information as given by depositor **Format**: see MINE format (Gams et al., J Gen Microbiol 1988;134:1667-1689). **Reference list** for country names: name based on country code file retrieved from: ftp://ftp.ripe.net/iso3166-countrycodes updated by the RIPE Network Co-ordination Centre, in co-ordination with the ISO 3166 Maintenance Agency at DIN Berlin. Exception for UK, USA, USSR, Netherlands, North Korea
Required for	RDS

Mutant

Field	Mutant
Description	type and parent of mutant if strain is a mutant strain. Applies only to a small number of strains.
Input process	Enter information as given by depositor
Required for	RDS

Sexual state

Field	Sexual state
Description	sexual condition of the strain. Does not apply to all strains
Input process	Format and reference list: see MINE format (Gams et al., J Gen Microbiol 1988;134:1667-1689).
Required for	RDS

Literature

Field	Literature
Description	reference to publication in which the strain is described. Especially recommended for type strains.
Input process	**new entry**: JournalTitle Year**;**Volume(issue)**:**beginning page#'**-**ending page# The journal title is abbreviated following international standard rules (ISSN abbreviations). The abbreviations are without dot. Authors and title of the article are not mentioned. **existing catalogues**: conversion by SRS following CABRI agreed standard This reference can be followed by the Pubmed ID enclosed within square brackets as follows: [PMID: 1234567], where '1234567' is the Pubmed ID of the paper.
Required for	RDS

Data Sets
Minimum Data Set, General Definition
Minimum data set for bacteria and archaea
Recommended data set for bacteria and archaea
Full data set
Data entry and validation
Accession number
Other culture collection numbers
Restrictions
Organism type
Name
Infrasubspecific names
Conditions for growth
Form of supply
Serovar
Isolated from
Geographic origin
Mutant

Status Other names History of deposit	Genotype Literature

Accession number

Field	Accession number
Description	collections' own accession number of the strain
Input process	consists of collection acronym followed by a number or alphanumeric identifier separated with a blank Make sure unique number is given to each strain
Required for	MDS

Other culture collection numbers

Field	Other culture collection numbers
Description	accession numbers of the strain in other collections
Input process	**New strain:** as given by depositor **Existing strain**: if strain is sent to another collection, get and add accession number used in that collection. It cannot be given if the strain is unique in the collection. numbers from individual researchers are also allowed Collection acronyms: WDCM registered collections and/or CABRI agreed list Numbers are separated by semicolon
Required for	MDS

Restrictions

Field	Restrictions
Description	value "no", or a text field describing the appropriate restrictions: hazard group, import/export regulations
Input process	Enter 'No', or actual restrictions, which may be linked to a file for conditions of delivery The text may contain codes which refer to text files explaining the restrictions in more detail.
Required for	MDS

Organism type

Field	Organism type
Description	contain value "Bacteria" or "Archaea".
Input process	Enter "B" for bacteria or "A" for Archaea
Required for	MDS

Name

Field	Name
Description	full scientific and most recent name of the strain: - Genus name and species epithet - Subspecies * - Pathovar * - Authors of the name - Year of valid publication or validation - Approbation of the name

Input process	Enter full scientific name as given by depositor and confirmed (or changed) by collection. Names of authors of the name, year of valid publication or validation and approbation are included after a comma. Values for approbation: AL = approved list, c.f.r. IJSB 1980 VL = validation list, in IJSB after 1980 VP = validly published, paper in IJSB after 1980 **Reference list**: DSMZ list of bacterial names
Required for	MDS

Infrasubspecific names

Field	Infrasubspecific names
Description	variety designation, epithet, authors and reference. Does not apply to all strains
Input process	Enter type and epithet of the variety as given by depositor and confirmed (or changed) by collection. Names of authors and reference are included. This field excludes the pathovar name and the serovar name, which are both infrasubspecific names but are to be entered into the NAME field Format: see MINE format (Stalpers et al., Systematic Applied Microbiol 1990; 13:92-103 http://www.urbanfischer.de/journals/sam/mic_biol.htm)
Required for	MDS

Status

Field	Status
Description	nomenclatural status of the strain (like "type", "neotype", etc.) Does not apply to all strains
Input process	Enter information on type and scientific name of organism, for which material from which strain was isolated is type material. Format: see MINE format (Stalpers et al., Systematic Applied Microbiol 1990;13:92-103).
Required for	MDS

Other names

Field	Other names
Description	names used for the strain in the past (not synonyms) that may still be in use. Same format as Name field. Does not apply to all strains.
Input process	Enter previous names used for the strain if they are no synonyms of the current name. Same format as Name field.
Required for	RDS

History of deposit

Field	History of deposit
Description	history of the element from depositor to isolator. - Name of depositor - Institute of depositor - Name of isolator - Institute of isolator - Names and institutes in between
Input process	Enter information as given by depositor Format: see MINE format (Stalpers et al., Systematic Applied Microbiol 1990;13:92-103).

Required for	MDS

Conditions for growth

Field	Conditions for growth
Description	Culture medium Atmospheric an light conditions Temperature conditions
Input process	Each collection has to provide its own list of culture media and recipes, medium provided for strain is linked to medium in this file. (http://www.dsmz.de/media/media.htm) Atmospheric and light conditions should only be given if they are special. Additional remarks on the cultivation like 'freshly prepared medium is necessary' or 'extended incubation time…' if necessary'
Required for	MDS

Form of supply

Field	Form of supply
Description	form in which the strain will be sent to the customer
Input process	Values: ‘Active', ‘Dried', 'Dried ice'
Required for	MDS

Serovar

Field	Serovar
Description	serovar name and author. Required for some medically important species. Does not apply to all strains
Input process	Format: see MINE format (Stalpers et al., Systematic Applied Microbiol 1990;13:92-103).
Required for	RDS

Isolated from

Field	Isolated from
Description	name of substrate or host from which the strain was isolated.
Input process	Enter information as given by depositor or as retrieved from public information sources Format: see MINE format (Stalpers et al., Systematic Applied Microbiol 1990;13:92-103).
Required for	RDS

Geographic origin

Field	Geographic origin
Description	name of country where strain originated from, followed by details on location.
Input process	Enter information as given by depositor or as retrieved from public information sources Format: see MINE format (Stalpers et al., Systematic Applied Microbiol 1990;13:92-103). **CABRI agreed list for country**: name based on country code file retrieved from: ftp://ftp.ripe.net/iso3166-countrycodes updated by the RIPE Network Co-ordination Centre, in co-ordination with the ISO 3166 Maintenance Agency at DIN Berlin. Exception for UK, USA, USSR, Netherlands, North Korea

Required for	RDS

Mutant

Field	Mutant
Description	type and parent of mutant if strain is a mutant strain. Applies only to a small number of strains.
Input process	Free text: enter information as given by depositor or as retrieved from public information sources
Required for	RDS

Genotype

Field	Genotype
Description	names of chromosomal markers of the strain. Especially recommended for strains of species for which many genetically modified strains exist. Does not apply to all strains
Input process	Format: see MINE format (Stalpers et al., Systematic Applied Microbiol 1990;13:92-103).
Required for	RDS

Literature

Field	Literature
Description	reference to publication in which the strain is described. Especially recommended for type strains
Input process	**new entry**: Journal Title Year**;**Volume(issue)**:**beginning page#**-**ending page# The journal title is abbreviated following international standard rules (ISSN abbreviations). The abbreviations are without dot. Authors and title of the article are not mentioned. **existing catalogues**: conversion by SRS following CABRI agreed standard This reference can be followed by the Pubmed ID enclosed within square brackets as follows: [PMID: 1234567], where '1234567' is the Pubmed ID of the paper
Required for	RDS

6.3.2 Data input & authentication for animal and human cell lines

Data entry and validation

1) Cell line name
2) Accession number
3) Brief description
4) Description and other remarks on the cell line
5) Depositor name and address
6) Reference paper
7) Morphology and growth characteristics
8) Culture conditions
9) Viruses
10) Properties
11) Passage
12) Release conditions/Restrictions
13) Hazard
14) Species validation tests

Cell line name

Field	Cell line name
Description	name given to the cell, originally name given by the depositor

Input process	**new entry for existing cells**:CABRI agreed list of cell names **new entry for new cells**: as given by depositor **existing catalogues**: retrieval by SRS should be possible using list of every possible spelling of a name.
Required for	MDS

Accession number

Field	Accession number
Description	collections' own accession number
Input process	automatically given, to avoid duplicates or absence of identifier
Note	a CABRI accession number is given, but only for internal purposes
Required for	MDS

Brief description

Field	Brief description
Description	brief listing of : - species - strain - tissue - tumour - pathology - transformed/transfected
Input process	Reference lists: - 'Dorland's Illustrated Medical Dictionary', 27th edition, Saunders -**strains** are in accordance with the 'Subsidiary of Harcourt Brace Jovanovich, Publishers', Academic Press: The Laboratory Rat, The Mouse in Biomedical Research, The Biology of the Laboratory Rabbit -**pathologies** are in accordance with the 'On-Line Mendelian Inheritance in Man (OMIM)',http://www.ncbi.nlm.nih.gov/Omim - CABRI agreed lists:species,tissues,tumours
Required for	MDS

Description and other remarks on the cell line

Field	Description and other remarks on the cell line
Description	contain ["verbose" listing] of: - species - strain - tissue - tumour - pathology - transformed/transfected - age - sex - derived from
Input process	Reference lists: - 'Dorland's Illustrated Medical Dictionary', 27th edition, Saunders -**strains** are in accordance with the 'Subsidiary of Harcourt Brace Jovanovich, Publishers', Academic Press: The Laboratory Rat, The Mouse in Biomedical Research, The Biology of the Laboratory Rabbit -**pathologies** are in accordance with the 'On-Line Mendelian Inheritance in Man (OMIM)',http://www.ncbi.nlm.nih.gov/Omim - CABRI agreed lists:species,tissues,tumours
Required for	MDS

Depositor name and address

Field	Depositor name and address

Description	depositor name address of the Institution
Input process	CABRI agreed list for country: name based on country code file retrieved from:ftp://ftp.ripe.net/iso3166-countrycodes updated by the RIPE Network Co-ordination Centre, in co-ordination with the ISO 3166 Maintenance Agency at DIN Berlin. Exception for UK, USA, USSR, Netherlands, North Korea
Required for	MDS

Reference paper

Field	Reference paper
Description	original paper [if available]
Input process	**new entry:** JournalTitle Year **;** Volume (issue)**:** beginning page# **-** ending page# The title is abbreviated following international standard rules (ISSN). Abbreviations are without dot. Authors and title of the article are not mentioned. **existing catalogues:** conversion by SRS following CABRI agreed standard This reference can be followed by the Pubmed ID enclosed within square brackets as follows: [PMID: 1234567], where '1234567' is the Pubmed ID of the paper
Required for	MDS

Morphology and growth characteristics

Field	Morphology and growth characteristics
Description	Morphology growth characteristics (monolayer, suspension, etc.)
Input process	Reference lists: -**morphology**: usingCABRI agreed list - ‘Dorland’s Illustrated Medical Dictionary’, 27th edition, Saunders
Required for	MDS

Culture conditions

Field	Culture conditions
Description	culture medium subculture routine incubation harvest
Input process	respect order of the items 1. culture medium, 2. subculture routine 3. incubation, 4. Harvest
Required for	MDS

Viruses

Field	Viruses
Description	absence/presence type of virus
Input process	In case of absence of data related to**Viruses**, input the following sentence:*"There is no evidence for the presence of infectious viruses"*. Reference lists: -CABRI agreed list - ‘Review of Medical Microbiology’, seventeenth ed., Appleton & Lange; - the WWW biological site ‘All the virology on the WWW’, (http://www.tulane.edu/~dmsander/garryfavweb.html)
Required for	MDS

Properties

Field	Properties
Description	Immunology receptors susceptibility to viruses
Input process	Free text
Requred for	MDS

Passage

Field	Passage
Description	passage no
Input process	when unknown, put ‘unknown’
Required for	*RDS*

Release conditions/Restrictions

Field	Release conditions/Restrictions
Description	release conditions/restrictions patents
Input process	**release conditions/restrictions**: enter "no", or a text field describing the appropriate restrictions: hazard group, import/export regulations, .. The text may contain codes which refer to text files explaining the restrictions in more detail. **patents**: EPO standard rules
Required for	MDS

Hazard

Field	Hazard
Description	hazard classification of the cell line [e.g. class II containment] EU related hazard information rules country-specific related hazard information rules
Input process	- classifications available on the CABRI Web site: see Guidelines for Quality Management - ACDP category list:http://www.doh.gov.uk/bioinfo.htm
Required for	MDS

Species validation tests

Field	Species validation tests
Description	Should contain tests and results of tests made to confirm species of cell lines
Input process	CABRI agreed list
Required for	*RDS*

6.3.3 Data entry and validation for plant cell lines

Scientific name
Family
Accession number
Strain designation
Depositor name and address
Culture conditions
Release conditions/Restrictions
Hazard
Additional information
Date of initiation

Scientific name

Field	Scientific name
Description	botanical name of initial plant
Input process	**Reference Lists:** - Taxonomic references: H. Melchior (ed.), "A. Engler's Syllabus der Planzenfamilien" Gebrüder Bornträger, Berlin 1964
Required for	MDS

Family

Field	Family
Description	Family whose belong the initial plant
Input process	**Reference Lists:** - Taxonomic references: H. Melchior (ed.), "A. Engler's Syllabus der Planzenfamilien" Gebrüder Bornträger, Berlin 1964
Required for	MDS

Accession number

Field	Accession number
Description	Collection Number given to the cell line
Input process	automatically given, to avoid duplicates or absence of identifier
Note	a CABRI accession number is given, but only for internal purposes
Required for	MDS

Strain designation

Field	Strain designation
Description	designations characterising the cell line
Input process	designations usually given by depositor or initiator
Required for	MDS

Depositor name and address

Field	Depositor name and address
Description	depositor name address of the Institution
Input process	CABRI agreed list for country: name based on country code file retrieved from: ftp://ftp.ripe.net/iso3166-countrycodes updated by the RIPE Network Co-ordination Centre, in co-ordination with the ISO 3166 Maintenance Agency at DIN Berlin. Exception for UK, USA, USSR, Netherlands, North Korea
Required for	MDS

Culture conditions

Field	Culture conditions
Description	- name of main cultivation medium (abbreviated) - recommended growth temperature (°C) - Humidity of growth chamber (%) - Light conditions (lux)

	- Light regime (hours)
Input process	Reference List: CABRI agreed list for abbreviated media: http://www.dsmz.de/media/media.htm
Required for	MDS

Release conditions/Restrictions

Field	Release conditions/Restrictions
Description	Release conditions/Restrictions patents
Input process	**release conditions/restrictions**: value "no", or a text field describing the appropriate restrictions: hazard group, import/export regulations, .. The text may contain codes which refer to text files explaining the restrictions in more detail. **patents:** EPO standard rules
Required for	MDS

Hazard/BioHazard

Field	Hazard/BioHazard
Description	hazard classification of the plant cell EU related hazard information rules country-specific related hazard information rules
Input process	See guidelines for Quality Management for Plant cell lines
Required for	MDS

Additional information

Field	Additional information
Description	any additional information about the cell line
Input process	Free text
Required for	MDS

Date of initiation

Field	Date of initiation
Description	date of culture initiation
Input process	Entered as dd.mm.yyyy
Required for	MDS

History

Field	History
Description	history of the cell line before arrival at the resource centre
Input process	
Required for	MDS

6.3.4 Data entry and validation for phages

Element name
Collection number
Element type
Other culture collection numbers
Restricted distribution
Literature

History of deposit
Host for propagation
Host used for propagation
Lysogenic
Cell surface receptor
Virus used for

Element name

Field	**Element name**
Description	Contains the name of the element
Input process	As given by the depositor
Required for	MDS

Collection number

Field	**Collection number**
Description	Contains the number of the element in the collection under which the element is distributed
Input process	collection acronym followed by a number
Required for	MDS

Element type

Field	Element type
Description	Contains element type.
Input process	**CABRI agreed list**: phage transposon minitransposon IS element
Required for	MDS

Other culture collection numbers

Field	Other culture collection numbers
Description	Numbers of the element under which it is distributed by other culture collections, incl. non-CABRI numbers if possible (e.g. ATCC, NCTC, plasmid collection of Vicente, …)
Input process	HGMP will be asked to generate the content of this field automatically, based on the name of the element. Clients must be warned that although the elements have the same name, the physical material may vary. Reference list: not necessary if automated
Required for	MDS

Restricted distribution

Field	**Restricted distribution**
Description	Restriction for the delivery and the use of the phage
Input process	enter "no", or a text field describing the appropriate restrictions: hazard group, import/export regulations, .. The text may contain codes which refer to text files explaining the restrictions in more detail.
Required for	MDS

Literature

Field	Literature
Description	Original publication in which the element is described.
Input process	**new entry**: JournalTitle Year**;**Volume(issue)**:**beginning page#'**-**'ending page# The journal title is abbreviated following international standard rules (ISSN abbreviations). The abbreviations are without dot. The title of the article is not mentioned. **existing catalogues**: conversion by SRS following CABRI agreed standard
Required for	MDS

History of deposit

Field	**History of deposit**
Description	Contains the history of the element from author to depositor, in reverse order. The name of the depositor must be entered obligatory; history when known.
Input process	Free text
Required for	MDS

Host for propagation

Field	**Host for propagation**
Description	Description of suitable host strains (on family, genus, species level), where possible completed with requirements for infection on phenotypic level (e.g. gram-negative bacteria, or *E. coli* LamB+)
Input process	Free text
Required for	MDS

Host used for propagation

Field	Host used for propagation
Description	Taxonomic designation and strain designation of the host organism used for propagation of the phage (e.g. E. coli MC1061)
Input process	Free text
Required for	MDS

Lysogenic

Field	**Lysogenic**
Description	Logical field: "yes" or "no". Enter "yes" when the infecting bacteriophage genome can become integrated into the bacterial host DNA, and is subsequently replicated and transmitted to progeny bacteria like any other chromosomal gene.
Input process	enter "yes" or "no",
Required for	MDS

Cell surface receptor

Field	**Cell surface receptor**
Description	Name of the cell surface receptor
Input process	Free text
Required for	RDS

Virus used for

Field	**Virus used for**

Description	Possible usage of the virus
Input process	Reference list: - PVT format (Vicente et al., World J Microbiol Biotechnol 1992;8:519-526: generalized transduction specialized transduction plasmid specificity phage typing other virus uses
Required for	MDS

6.3.5 Data entry and validation for plasmids

Name
Collection number
Type
Class
Other culture collection numbers
Host for distribution
Medium
Restricted distribution
Literature
History of deposit
Selectable phenotype
Replicon
Host range
Properties and applications

Name

Field	Name
Description	Name of the element
Input process	Name given by the depositor or as published
Required for	MDS

Collection number

Field	Collection number
Description	Number of the element in the own collection under which it is distributed
Input process	Collection acronym followed by a number separated by space
Required for	MDS

Type

Field	Type
Description	Element Type
Input process	**Reference list**: plasmid phasmid cosmid shuttle vector transposon minitransposon IS element
Required for	MDS

Class

Field	Class
Description	Origin of the element

Input process	**Reference list**: non-recombinant, recombinant
Required for	MDS

Other culture collection numbers (*)

Field	Other culture collection numbers
Description	Numbers of the element under which it is distributed by other culture collections.
Input process	This field is automatically generated. (*) If the content of this field cannot automatically be generated at the main site it will be removed from the MDS for plasmids. This field does not apply to all strains.
Required for	MDS

Host for distribution

Field	Host for distribution
Description	Taxonomic designation and strain designation of the host organism in which the element is distributed
Input process	e.g. Escherichia coli K-12 MC1061 Reference list: CABRI agreed list
Required for	MDS

Medium

Field	Medium
Description	Name or collection number of the medium on which the host organism is preferably grown
Input process	Free text which may be linked to a text file describing the recipe for the medium
Required for	MDS

Restricted distribution

Field	Restricted distribution
Description	Restrictions for distribution of the element
Input process	Restrictions for distribution of the element (hazard group, "for research purposes only", patent info., import/export regulations, etc.
Required for	MDS

Literature

Field	Literature
Description	Original publication in which the element is described.
Input process	**new entry**: Author, JournalTitle Year; Volume (issue): beginning page#-ending page# The journal title is abbreviated following international standard rules. The abbreviations are without dot. Title of the article is not mentioned. **existing entries**: conversion by SRS following CABRI agreed standard. This field does not apply to all strains. This reference can be followed by the Pubmed ID enclosed within square brackets as follows: [PMID: 1234567], where '1234567' is the Pubmed ID of the paper
Required for	MDS

History of deposit

Field	History of deposit

Description	History of the element from author to depositor, in reverse chronology (depositor< … < author).
Input process	The name of the depositor must be entered obligatory; history when known. This field does not apply to all strains: for plasmids deposited before december 1993 (cfr. CBD convention), this field may lack information.
Required for	MDS

Selectable phenotype

Field	Selectable phenotype
Description	Feature encoded by the element enforcing its host strain to maintain the element in order to survive under appropriate cultivation conditions
Input process	Reference list: CABRI agreed list
Required for	MDS

Replicon

Field	Replicon
Description	Names of replicon elements
Input process	Reference list: CABRI agreed list This field does not apply to all strains.
Required for	MDS

Host range

Field	Host range
Description	Host in which the element can be propagated and/or used and, when appicable, the peculiar features that are necessary for making use of the special characteristics of the element.
Input process	Free text based on the reference list **VECAP3 list** in Vicente et al., World J Microbiol Biotechnol 1992;8:519-526 Ex.: Escherichia coli preferably containing the lacI^{q} gene This field does not apply to all strains.
Required for	MDS

Properties and applications

Field	Properties and applications
Description	Description of the properties of the element, as well as its application possibilities
Input process	Reference list: CABRI agreed list based on the VECAP4, VECAP5 and VECAP6 list in Vicente et al., World J Microbiol Biotechnol 1992;8:519-526
Required for	MDS

6.3.6 Data entry and validation for viruses

Accession number
Virus name
Virus Name Abbreviation
Former Name
Genus
Pathotype, Serotype, Strain
Original Host
Geographic Origin
Isolate History
Reference Isolate
Quarantine Regulations
Remarks

Accession Number

Field	Accession Number
Description	Accession number assigned by the collection
Input process	unique identifier
Required for	MDS

Virus Name

Field	Virus Name
Description	Valid species name of the virus
Input process	Should be entered according to the: *Sixth Report of the International Committee on Taxonomy of Viruses* (ICTV), 1995: http://www.ncbi.nlm.nih.gov/ICTV/ Names that have not yet been approved by ICTV are printed between quotation marks.
Required for	MDS

Virus Name Abbreviation

Field	Virus Name Abbreviation
Description	Species name abbreviation
Input process	abbreviation entered according to the ICTV, 1995: http://www.ncbi.nlm.nih.gov/ICTV/ abbreviations that have not yet been approved by ICTV are printed between quotation marks.
Required for	MDS

Former Name

Field	Former Name
Description	Former name of the isolate, e.g., in case a patho- or serotype has been assigned to another species
Input process	Former name entered according to the ICTV: http://www.ncbi.nlm.nih.gov/ICTV/
Required for	MDS

Genus

Field	Genus
Description	Genus to which the virus species has been assigned
Input process	Genus name entered according to the ICTV: http://www.ncbi.nlm.nih.gov/ICTV/
Required for	MDS

Pathotype, Serotype, Strain

Field	Pathotype_Serotype_Strain
Description	"Pathotype: ", "Serotype: " or "Strain: " followed by pathotype, serotype or strain designation.
Input process	
Required for	MDS

Original Host Plant

Field	Original_Host
Description	should contain: - Original host plant

	- Subspecies - variety - cultivar or breeding line - vernacular name
Input process	**Original host plant**: - species (latin binominal), - subspecies (latin), - variety (latin) - cultivar or breeding line (original designation), **vernacular name** (English or original designation). The vernacular name is listed only in special cases, e.g., if the vernacular name given by the depositor cannot be unequivocally assigned to a plant species.
Required for	MDS

Geographic Origin

Field	Geographic Origin
Description	Country, and, if available, more detailed information on the geographic position where the original host plant was found.
Input process	Free text
Required for	MDS

Isolate History

Field	Isolate History
Description	Name of the scientist, who isolated the specimen and of persons or organisations that kept it before accessioning.
Input process	After each name (person or organisation), the respective isolate designation is given between single quotation marks.
Required for	MDS

Reference Isolate

Field	Reference Isolate
Description	Literature on the virus isolate
Input process	**new entry**: JournalTitle Year;Volume(issue):beginning page#-ending page# The journal title is abbreviated following international standard rules. The abbreviations are without dot. Authors and title of the article are not mentioned. **existing catalogues**: conversion by SRS following CABRI agreed
Required for	MDS

Quarantine Regulations

Field	Quarantine Regulations
Description	should contain: "yes" if EU quarantine regulations for the virus species are known.
Input process	Yes/no
Required for	MDS

Remarks

Field	**Remarks**
Description	Special features of the isolate, e.g., virulence
Input process	Free text

Required for	MDS

6.3.7 Data entry and validation for genomic libraries

Library Name
Organism
Type of Library
Vector
Insert Size
Library coverage
Vector
Library coverage
MTA required

Library Name

Field	Library Name
Description	library Identifier e.g. WI/MIT, RCPI1
Input Process	Ensure Every Library has Unique identifier, New Library as given by depositor.
Required for	MDS

Organism

<table>
<tr><td>Field</td><td>Organism</td></tr>
<tr><td>Description</td><td>The Organism represented in the Library</td></tr>
<tr><td>Input Process</td><td>CABRI agreed list (Note: may have multiple species):
<table>
<tr><th>Organism</th><th>Synonym</th></tr>
<tr><td>Mouse</td><td>Murine, Mus Musculus</td></tr>
<tr><td>Human</td><td>Homo Sapiens</td></tr>
<tr><td>Rat</td><td></td></tr>
<tr><td>Fugu</td><td></td></tr>
<tr><td>Drosophila</td><td></td></tr>
<tr><td>Dog</td><td>Canine</td></tr>
</table></td></tr>
<tr><td>Required for</td><td>MDS</td></tr>
</table>

Type of Library

Field	Type
Description	Type of Library
Input Process	CABRI agreed list: PAC BAC YAC PI cDNA
Required for	MDS

Vector

Field	Vector
Description	Vector name , more specific than type as will be specific about vector e.g. pCYPAC, Lawrist4

Input Process	Supplied by depositor, may have multiple vectors per library
Required for	MDS

Insert Size

Field	Insert Size
Description	Average size of inserts in vector in Kb
Input Process	Numeric entry
Required for	MDS

Library coverage

Field	Library coverage
Description	Degree of coverage
Input Process	Numeric entry
Required for	MDS

MTA required

Field	MTA required
Description	Is a material Transfer Agreement required
Input Process	"Yes" or "No"
Required for	MDS

6.3.8 Data entry and validation for hybridomas

Hybridoma name
Accession number
Brief description
Description
Depositor
Reference paper
Morphology and growth characteristics
Culture conditions
Release conditions/Restrictions
Hazard

Hybridoma name

Field	Hybridoma name
Description	a common agreed list of hybridoma names, considering the orthographic differences in the different catalogues
Input process	**new entry for existing hybridomas**: using common agreed list of hybridoma names **new entry for new hybridomas**: as given by depositor **existing catalogues**: retrieval by SRS should be possible using list of every possible spelling of a name.
Required for	MDS

Accession number

Field	Accession number
Description	collections’ own accession No or hybridoma code
Input process	automatically given, to avoid duplicates or absence of identifier
Note	a CABRI accession number is given, but only for internal purposes

Required for	MDS

Brief description

Field	Brief description
Description	["brief" listing]: - species - strain - antibody specificity
Input process	**Vocabulary must be checked using :** ‘Dorland’s Illustrated Medical Dictionary’, 27th edition, Saunders **strains** are in accordance with the ‘Subsidiary of Harcourt Brace Jovanovich, Publishers’, Academic Press: The Laboratory Rat, The Mouse in Biomedical Research, The Biology of the Laboratory Rabbit
Required for	MDS

Description and other remarks on the hybridoma

Field	Description and other remarks on the hybridoma
Description	["verbose" listing]: - species - strain - antibody specificity - donor [immunised animal] - antibody produced - cross-reaction - fusion partner
Input process	**Vocabulary must be checked using :** ‘Dorland’s Illustrated Medical Dictionary’, 27th edition, Saunders **strains** are in accordance with the ‘Subsidiary of Harcourt Brace Jovanovich, Publishers’, Academic Press: The Laboratory Rat, The Mouse in Biomedical Research, The Biology of the Laboratory Rabbit Additional information can be found in "ICDB Immunoclone Database, coding manual", contact CERDIC.
Required for	MDS

Depositor name and address

Field	Depositor name and address
Description	- depositor name - address of the Institution
Input process	CABRI agreed list for country: name based on country code file retrieved from: ftp://ftp.ripe.net/iso3166-countrycodes updated by the RIPE Network Co-ordination Centre, in co-ordination with the ISO 3166 Maintenance Agency at DIN Berlin. Exception for UK, USA, USSR, Netherlands, North Korea
Required for	MDS

Reference paper

Field	Reference paper
Description	Original paper [if available]
Input process	**new entry**: JournalTitle Year**;**Volume**(**issue**):**beginning page#**-**ending page# The title is abbreviated following international standard rules. The abbreviations are without dot. Authors and title of the article are not mentioned. **existing catalogues**: conversion by SRS following CABRI agreed standard
Required for	MDS

Morphology and growth characteristics

Field	Morphology and growth characteristics
Description	morphology growth characteristics (monolayer, suspension, etc.)
Input process	**Vocabulary must be checked using :** **morphology:** 'Dorland's Illustrated Medical Dictionary', 27th edition, Saunders
Required for	MDS

Culture conditions

Field	Culture conditions
Description	- culture medium - subculture routine - incubation - harvest
Input process	respect order of the items 1.culture medium, 2. subculture routine 3. incubation, 4. Harvest
Required for	MDS

Release conditions/Restrictions

Field	Release conditions/Restrictions
Description	- release conditions/restrictions - patents
Input process	**release conditions/restrictions**: enter "no", or a text field describing the appropriate restrictions: hazard group, import/export regulations, .. The text may contain codes which refer to text files explaining the restrictions in more detail. **patents**: EPO standard rules
Required for	MDS

Hazard

Field	Hazard
Description	- hazard classification of the hybridoma - EU related hazard information rules - country-specific related hazard information rules
Input process	Classification to be found on the guidelines for quality management
Required for	MDS

6.4 Field labels

In order to harmonise and optimise searches in several catalogues at the same time, it has been agreed to define labels that must specify the information contents of each field. Labels are common for catalogues of the same biological materials.

All Collections, including the centres wishing to enter CABRI, MUST adhere to the following lists for the MDSs and the RDSs. IF and ONLY IF new collection centres plan to put in their own FDS an information that is already present in other FDSs THEN they are invited to adopt the existing fields.

In the following list, for each field of the Full Data Set (FDS) the catalogues that use it are also listed.

6.4.1 Animal and human cell lines

SET Field name
MDS Accession_number
MDS Cell_line_name
MDS Brief_description
MDS Description
MDS Depositor
MDS Bibliographic_references
MDS Morphology
MDS Culture_conditions
MDS Viruses
MDS Properties
MDS Release_conditions
MDS Hazard
RDS Passage_number
RDS Species_validation
FDS Tumorigenicity
FDS Karyology
FDS Freezing_medium
FDS Sterility
FDS Validation_assays
FDS Further_bibliography
FDS Comments
FDS Storage
FDS Doubling_time
FDS Mycoplasma
FDS Fingerprint
FDS Cytogenetics
FDS Karyotype
FDS Comments
FDS Research_council_deposit
FDS BIOMED_1

6.4.2 Hybridomas

Order and display field labels

SET Field name
MDS Accession_number
MDS Hybridoma_name
MDS Brief_description
MDS Description
MDS Depositor
MDS Reference_paper
MDS Morphology_and_growth
MDS Culture_conditions
MDS Distribution
MDS Hazard
FDS Assays
FDS Applications

FDS Research_council_deposit
FDS Amend_date
FDS Catalogue_version

6.4.3 Fungi

Order and display field labels

SET Field name
MDS Strain_number
MDS Other_collection_numbers
MDS Name
RDS Race
MDS Organism_type
MDS Restrictions
MDS Status
MDS History
RDS Misapplied_names
RDS Substrate
RDS Geographic_origin
RDS Sexual_state
RDS Mutant
FDS Genotype
FDS Pathogenicity
FDS Enzyme_production
FDS Metabolite_production
RDS Applications
RDS Literature
FDS Remarks
MDS Conditions_for_growth
MDS Form_of_supply
FDS Price_code

6.4.4 Yeasts

Order and display field labels

SET Field name
MDS Strain_number
MDS Other_collection_numbers
MDS Name
MDS Organism_type
MDS Restrictions
MDS Status
MDS History
RDS Geographic_origin
RDS Sexual_state
RDS Mutant
FDS Genotype
FDS Pathogenicity
FDS Enzyme_production
FDS Metabolite_production
RDS Applications
RDS Literature

FDS Remarks
MDS Conditions_for_growth
MDS Form_of_supply
FDS Price_code

6.4.5 Bacteria & Archaea

Order and display field labels

SET Field name
MDS Strain_number
MDS Other_collection_numbers
MDS Restrictions
MDS Organism_type
MDS Name
MDS Infrasubspecific_names
MDS Status
MDS History
MDS Conditions_for_growth
MDS Form_of_supply
RDS Serovar
RDS Other_names
RDS Isolated_from
RDS Geographic_origin
FDS Sexual_state
RDS Mutant
RDS Genotype
RDS Literature
FDS Pathogenicity
FDS Enzyme_production
FDS Metabolite_production
FDS Applications
FDS Catalogue_entry
FDS Remarks
FDS Price_code
FDS Plasmids

6.4.6 Plasmids

Order and display field labels

SET Field name
MDS Collection_number
MDS Name
MDS Other_culture_collection_numbers
MDS Type
MDS Class
FDS Constructed_from (DSMZ_PLASMID, NCCB_PLASMID)
MDS Literature
MDS History_of_deposit
MDS Restricted_distribution
MDS Host_for_distribution
MDS Medium
MDS Selectable_phenotype

MDS Replicon
FDS Incompatibility_group (DSMZ_PLASMID, NCCB_PLASMID)
FDS Transfer_ability (DSMZ_PLASMID, NCCB_PLASMID)
MDS Host_range
MDS Properties_and_applications
FDS Helper
FDS Copy_number
FDS Molecular_weight
FDS Cloned_gene
FDS Transposable_element
FDS Promoter
FDS Ribosome_binding_site
FDS Start_codon
FDS Terminator
FDS Further_information*
FDS Restriction_sites
FDS Sequence_detail
FDS Price code

* Description of the field "Further_information": remarks on propagation and/or on properties and/or on history, other name(s), etc.

6.4.7 Plant cell lines

Order and display field labels

SET Field name
MDS Collection_number
MDS Scientific_name
MDS Family
MDS Strain_designation
MDS Medium_1
MDS Temperature
MDS Humidity
MDS Light_conditions
MDS Light_regime
MDS Hazard
MDS Additional_information
MDS Depositor
MDS History
MDS Initiation_date
MDS Restrictions
FDS Initiation_material
FDS Accession_date

6.4.8 Phages

SET Field name
MDS Collection_number
MDS Name
MDS Type
MDS Other_collection_numbers
MDS Restricted_distribution
MDS Literature

MDS History_of_deposit
MDS Host_for_propagation
MDS Host_used_for_propagation
MDS Lysogenicity
MDS Virus_used_for
RDS Cell surface receptor
?DS Applications
?DS Host_range
FDS Medium
FDS Class
FDS Constructed_from
FDS Host_for_selection
FDS Viral_morphology
FDS Plaque_forming_phage
FDS Clear_plaque_morphology
FDS Plaque_description
FDS Particle_size
FDS Minimum_packaging_capacity
FDS Maximum_packaging_capacity
FDS Properties_and_applications
FDS Remarks_on_applications
FDS Cloned_gene
FDS Transposable_element
FDS Start_codon
FDS Further_information
FDS Restriction_sites
FDS Price_code

6.4.9 Plant viruses

SET Field Name
MDS Accession_number
MDS Virus_name
MDS Virus_name_abbreviation
MDS Former_name
MDS Genus
MDS Pathotype_serotype_strain
MDS Original_host_plant
MDS Geographic_origin
MDS Isolate_history
MDS Reference
MDS Quarantine_regulations
MDS Remarks

7 Bibliography

BioMOBY***:*** http://biomoby.org/

GBIF Standards:
Data exchange formats and communication protocols
ABCD Access to Biological Collection Data - a joint CODATA and TDWG initiative
BioCASE The project and protocol for exchanging ABCD data
Darwin Core A simple exchange format for specimen and observation data
DiGIR Distributed Generic Information Retrieval
Dublin Core Metadata standard for publishing
GGF Global Grid Forum
Invasive Species Profile GISIN invites you to comment on the invasive alien species profile draft schema.
LSID Life Sciences Identifier is a new naming standard and data-access protocol under development. Send comments on this link to DADI.gbif@ig.circa.gbif.net with subject LSID.
TDWG International Working Group on Taxonomic Databases
Unified DiGIR / BioCASe Protocol Wiki
Data Quality, Data Cleaning and Data Use - http://en.wikipedia.org/wiki/Secure_Sockets_Layer
SOAP (Simple Object Access Protocol)
SOAP Version 1.2 Part0: Primer - http://www.w3.org/TR/2003/REC-soap12-part0-20030624/
SOAP Version 1.2 Part1: Messaging Framework - http://www.w3.org/TR/2003/REC-soap12-part1-20030624/
SOAP Version 1.2 Part2: Adjuncts - http://www.w3.org/TR/2003/REC-soap12-part2-20030624/
SOAP Version 1.2 Specification Assertions and Test Collection - http://www.w3.org/TR/2003/REC-soap12-testcollection-20030624/
XML –Extensible Markup Language - http://www.w3.org/XML/

Useful Websites
Biocase - A Biological Collection Access Service for Europe - http://www.biocase.org/default.shtml
DBJ/EMBL/GenBank - DNA Data Bank of Japan, Mishima, Japan., EMBL Nucleotide Sequence Database, Cambridge, UK and GenBank, NCBI, Bethesda, MD, USA. http://www.ebi.ac.uk/embl/Documentation/FT_definitions/feature_table.html
GBIF – Global Biodiversity Information Facility - http://www.gbif.org/
OASIS - Organization for the Advancement of Structured Information Standards http://www.oasis-open.org/home/index.php
W3C – World Wide Web Consortium - http://www.w3.org/
CABRI – Common Access to Biological Resources and Information – http://www.cabri.org/

8 ANNEXES

8.1 Tools, Protocols and Schemes

Whatever the internal information system is, an mBRC must present data in accordance with the CABRI standard (see below). The most appropriate combined package of tools, protocols and schemes for information linkage and exchange of the Network and mBRCs should be identified to adopt at Information network. The CABRI standard provides the existing models of tools, protocols and schemes which are recommended.

8.1.1 Unique Identifier (UI)

The UI should fully utilise existing world directories where available, e.g. the World Directory of Collections and Cultures of Microorganisms (the database of CCINFO). The UI will be utilised not only by the Network but also other global information networks such as GBIF and International Nucleotide Sequence Database (DDBJ/EMBL/GenBank). The UI for the holding will have a triplet structure consisting of the mBRC code, the collection code in the mBRC and the unique number of the specimen in the collection

8.1.2 XML (Extensible Mark-up Language), XML Schema, the Extensible Stylesheet Language Family (XSL)

Extensible Markup Language (XML) is a simple, very flexible text format derived from SGML (ISO 8879). Originally designed to meet the challenges of large-scale electronic publishing, XML is also playing an increasingly important role in the exchange of a wide variety of data on the Web and elsewhere.

XML Schemas express shared vocabularies and allow machines to carry out rules made by people. They provide a means for defining the structure, content and semantics of XML documents in more detail.

XSL is a language for expressing stylesheets. Given a class of arbitrarily structured XML documents or data files, designers use an XSL stylesheet to express their intentions about how that structured content should be presented; that is, how the source content should be styled, laid out, and paginated onto some presentation medium, such as a window in a Web browser or a hand-held device, or a set of physical pages in a catalogue, report, pamphlet, or book.

8.1.3 XML-Security

XML-Security is a system that enables user access control coupled with digital rights management facilities. This is particularly useful in the context of information network since it allows tracking of data thereby enabling the preservation of IPR.

8.1.4 CAS (Common Access System) as an Interface

This will provide a wrapper to enable data from the heterogeneous underlying databases of member mBRCs to be accessed in a single uniform manner. Tools to produce these wrappers will need to be produced by the central network. Web services is a comparatively simple technology for the wrapping. GRID technology is equipped with authentication of participants but it is a comparatively complex system.

8.1.5 DarwinCore and ABCD (Access to Biological Collection Data)

GBIF uses both DarwinCore and ABCD schema for the data format and will integrate them into a unified scheme. ABCD and BioCASE will be provided as an option of schema and protocol for Information network.

8.1.6 SOAP (Simple Object Access Protocol)

SOAP provides the definition of the XML-based information which can be used for exchanging structured and typed information between peers in a decentralised, distributed environment.

8.1.7 Data Structure

The assumption is that the data will be extracted from the underlying databases by the CAS as XML in common with other similar networks such as GBIF.
Standards for data format (XML), metadata (XML Schema), data transformation (XSL) and data transport protocol, e.g. SOAP (Simple Object Access Protocol), could be derived or directly used from existing tools (e.g. based on W3C standards, from BioMOBY and from GBIF).
The ABCD (Access to Biological Collection Data) being produced as a result of collaboration between CODATA and TDWG. This has a sufficiently rich structure to allow its use in the context of information network and has the advantage that collections forming part of more than one network, could use the same general format for data exchange and retrieval. In the meantime, DarwinCore is more popular in GBIF data providers as of September 2005.

8.1.8 Data Security and Client Authentication

While the well known methods of use rid/password and SSL (secure sockets layer) can be used for XML documents, they are not ideal as they do not take into account the structure of XML data. In addition, while SSL ensures data security during the transmission, on arrival the data may be left unsecured on an unprotected server. The W3C consortium and the OASIS group are currently producing a series of XML-based security standards under the generic name XML-Security.
XML-Security allows the implementation of user access control coupled with digital rights management facilities. This is particularly useful in the context of information network since it allows tracking of data thereby enabling the preservation of IPR.
XML-Security is 'aware' of the deeply structured nature of XML data and it is feasible to make portions of particular structured data only accessible to particular classes of user. For example, in a BRC of micro-organisms, certain strains of an accession, which may be of interest in terms of biosecurity, could be restricted to only a high category of user and remain opaque to others.

8.1.9 Domain-specific Data Structure

i) Micro-organism domain of BRCs
Minimum data set for Micro-organism domain
XML schema for Micro-organism domain

9 Conclusion

The EMbaRC project did not include an element for creating a specific information system to give access to partners' catalogue data so existing systems were explored. The straininfo.net system provides access to culture collection "passport" data and facilitates linkages between lines of the same strain between culture collections and facilitates links to literature and sequence data for each strain (www.straininfo.net).

Dawyndt, P.; Vancanneyt, M.; De Meyer, H.; Swings, J., Knowledge accumulation and resolution of data inconsistencies during the integration of microbial information sources, IEEE Transactions on Knowledge and Data Engineering, vol.17, no.8, pp. 1111-1126, Aug. 2005

This system was considered as a useful tool but did not give access to the EMbaRC partner catalogue data. However, all EMbaRC partners were encouraged to supply their data to this project.

A second existing data system explored was the Common Access to Biological Resources and Information project (CABRI) www.cabri.org. Again all EMbaRC partners were encouraged to join this electronic catalogue. CABRI is an ideal starting point for a gateway to those collections adopting quality standards and could provide an initial data system for EMbaRC and the larger projects of MIRRI and the GBRCN. All collections accepted into CABRI must undergo an audit from the CABRI Technical Committee and then submit their data as specified on the CABRI website. The process for this is explained at http://www.cabri.org/guidelines/procedures/procedures_manual.html#join. The guidelines for catalogue production can be found at http://www.cabri.org/guidelines/catalogue/CPcover.html have been used as a basis for this data management and exchange standard. Only three of the EMbaRC partners are not present, MUM, CECT and CIRM, INRA. Unfortunately, the audit teams need to be funded and again the EMbaRC project had not allowed for this in its budget. Each partner collection must meet the CABRI guidelines, however, all that meet OECD best practice linked to ISO 90001 series certification should have no problems in successfully getting through the CABRI audit.

A third opportunity to link EMbaRC data is through the World Data Centre for Microorganisms (WDCM) new initiative to establish a global microbial resources database and again all EMbaRC partners have been encouraged to join (www.wfcc.info/home/wdcm/).

Regardless of the option taken these *Guidelines for optimal formatting/annotation of data related to the biological materials* can form the basis for data management and exchange.

Significance of this deliverable

EMbaRC project did not have the aim to develop new system but had the aim to study the existing systems. It has been performed via this deliverable. All EMbaRC partners were encouraged to link their data to straininfo, CABRI and WDCM.